On Sharing Viral Video Over an Ad Hoc Wireless Network

Yi-Ting Chen, Constantine Caramanis and Sanjay Shakkottai
Department of Electrical and Computer Engineering
The University of Texas at Austin
Email: {yiting.chen, caramanis, shakkott}@mail.utexas.edu

Abstract—We consider the problem of broadcasting a viral video (a large file) over an ad hoc wireless network (e.g., students in a campus). Many smartphones are GPS enabled, and equipped with peer-to-peer (ad hoc) transmission mode, allowing them to wirelessly exchange files over short distances rather than use the carrier's WAN. The demand for the file however is transmitted through the social network (e.g., a YouTube link posted on Facebook).

To address this coupled-network problem (demand on the social network; bandwidth on the wireless network) where the two networks have different topologies, we propose a file dissemination algorithm. In our scheme, users query their social network to find geographically nearby friends that have the desired file, and utilize the underlying ad hoc network to route the data via multi-hop transmissions. We show that for many popular models for social networks, the file dissemination time scales sublinearly with n, the number of users, compared to the linear scaling required if each user who wants the file must download it from the carrier's WAN.

I. INTRODUCTION

The proliferation of mobile devices that can stream video (laptops, smartphones, tablets) has marked a dramatic increase in demand for streaming video. At the same time, content generation and dissemination has become dramatically easier – most phones have installed video-cameras, and knowledge of a video can spread extremely rapidly to vast numbers of people, through social networks including e-mail, Facebook, Twitter, and the like. As deployed capacity approaches saturation, we need new transmission architectures to guarantee our wireless networks continue to deliver traffic effectively and efficiently.

This paper addresses precisely this problem. More specifically: we consider the simple, yet increasingly common setting, where a user (e.g., a student on a college campus) generates a large file (a short video, for example) and wants to spread it to her social network – her friends, their friends, and so on. In the current paradigm, the file creator uploads the file to a central server (e.g., YouTube) and then spreads word of its existence through Facebook, Twitter, etc. Upon learning of the file's existence, interested (we call them "eager") users then download the file from the server, using their provider's wide area network (WAN). Since the WAN has bounded bandwidth, the file dissemination time will necessarily scale linearly in the number of users who ultimately receive the file. Particularly in a dense setting like a college campus, this inherently limited centralized scheme for file dissemination

may be highly suboptimal. The central question in this paper is: how much better can we do?

Increasingly, smartphones and similar technology, are equipped with both GPS and peer-to-peer transmission modes. In dense environments, this opens the possibility of forming a wireless ad hoc network in which users communicate with each other through several hops of short distance transmissions. As shown in Gupta and Kumar's seminal work [1], the spatial capacity of a wireless ad hoc network scales as $\sqrt{n}$ – a sharp contrast to the fixed capacity of a WAN. While this scaling spatial capacity of ad hoc networks provides a potential way forward, naive implementation presents severe problems that may leave us worse off than the currently implemented WAN solution. We may have severe congestion caused by subsets of users getting a high number of requests, hence resulting in hot-spots in the network. This will occur, for instance, if users request the file from neighbors on their social network, as most social networks exhibit the presence of super-nodes with very high degree. This is particularly true in the broadcast setting we have here, when we expect there to be such hot spots, which can potentially reduce network capacity by a significant factor [17].

A. Main contributions

In this paper we propose a simple and distributed file dissemination algorithm that takes advantage of two main ideas: (i) knowledge of the file spreads quickly because of the structure of the social network – we can use the same to manage file dissemination; (ii) in dense settings where ad hoc networks make sense, exploiting geographic proximity can provide additional benefits. With these ideas in mind, we devise a file dissemination algorithm that works by passing messages through the social network, and requires limited communication and computation overhead. In particular, the main features of our algorithm are as follows:

1) Load balancing: users receiving a large amount of requests distribute them to nearby users on the social network, in such a way that we can guarantee no user has to serve more than six other users. Our algorithm achieves $\sqrt{n}$-scaling with the number of users receiving the file – sublinear, in sharp contrast to the linear scaling required in the WAN file dissemination architecture.

2) Exploiting geographic proximity: We extend our load-balancing algorithm to exploit geographic proximity.

Because of the structure of the social network, we show that by searching a few hops deeper in their social network, most users are able to download the file from another user at close range. This idea allows us to further reduce the scaling below $\sqrt{n}$, depending on the depth of the social-network a user may search.

3) Social Networks: We analyze our algorithm on popular models for social networks (power law graphs). We show that the file dissemination time scales sublinearly with n for a broad range of social-network parameters. In addition, we show that the performance of our algorithm is comparable to the best possible dissemination time of any algorithm – even those not constrained by communication or computation time.

B. Related work

Single piece file dissemination problems were first studied in [18][19]. In [20]-[22], they provide analytic results for multi-piece file dissemination problems. Other topics related to influence spreading, epidemics, and content distribution in social networks can be found, for example, in [23]-[25] and references therein.

Multi-hop transmission in a wireless network has been studied extensively since Gupta and Kumar's seminal work [1]. Subsequently, [3] provides a simple proof and [2] closes the gap of $1/\sqrt{\log(n)}$. Multicast and broadcast capacities are considered in, e.g., [11]-[13]. On the other hand, [14]-[17] use randomized schemes to balance the traffic load and achieve throughput optimal routing.

C. Paper organization

We introduce the system model in Section II. In Section III, we present our algorithm and main results. Some lemmas regarding random placement and random graphs are included in Section IV. We analyze the performance of our algorithm in Section V. Conclusions are provided in Section VI. The proofs of various lemmas and theorems in Section III and Section IV can be found in Appendix A and Appendix B.

II. SYSTEM DESCRIPTION

In this section we describe the basic system model, including the model for the wireless network and the placement of the nodes, and the model for the social network.

A. Random wireless network and Gaussian channel model

We model our network as n static nodes, placed independently and uniformly on a square of width $\sqrt{n}$. Thus the (expected) density of the network stays constant. Each node has a transmitter and a receiver. All nodes can communicate with each other with fixed power P. The interference model is described by a Gaussian channel model defined below.

Definition 1: (Gaussian channel model) Index nodes by 1, 2, ..., n. Let x_i be the location of node i. Let $\mathcal{A}$ be the set of active transmitters at this time instant. The transmission rate

$R(x_i, x_j)$ from node i to node j is

$$R(x_i, x_j) = \log\left(1 + \frac{P\ell(x_i, x_j)}{N_0 + \sum_{k \in \mathcal{A}\setminus\{i\}} P\ell(x_k, x_j)}\right). \quad (1)$$

Here, $\ell(x, y)$ represents the power attenuation function between points x and y on the square, and is given by

$$\ell(x_i, x_j) = \min\left\{1, \frac{e^{-\gamma||x_i - x_j||}}{||x_i - x_j||^\alpha}\right\} \quad (2)$$

where as usual, $||x - y||$ is the Euclidean distance between x and y.

In this paper, we consider either $\gamma > 0$ or $\gamma = 0$ and $\alpha > 2$.

B. Model for social networks

As we identify users with their devices (e.g. cell phones/ PDA), the n nodes in the wireless network also form a social network. A social network is described as a graph $G = (V, E)$ where V is the set of nodes with cardinality n and E is the set of edges. Two nodes are joined by an edge if (and only if) the corresponding users are friends in the social network. The distance between two nodes x and y on the social-graph G is the minimum number of hops between x and y in the social network. Thus a node's neighbors are the nodes one hop away on the social graph, and its k-neighborhood are the nodes within k hops away on the social graph. A key property we exploit is that distance between two nodes on the social network is generally unrelated to geographic distance between the corresponding users in the wireless network.

Empirical studies of many social (and other) networks have shown them to satisfy so-called power law graph structure, including many collaboration networks, but also the Internet and many communication networks (see e.g. [7] [8] [9] [10]). As a consequence, power law graphs (which we define below) are a popular choice for modeling social networks.

A graph G is called a power law graph with parameter β if the number of nodes with degree k is proportional to $k^{-\beta}$. We will consider social networks generated by random power law graphs [4]. These random graphs satisfy an important property: each node has only small number of neighbors, i.e., small degree (small relative to the size of the overall network) while the diameter of the random graph (the maximum number of hops between the vast majority of the nodes) is still small, with overwhelming probability. This property is consistent with properties of most social networks, and in particular, with the famous observation known as the small world phenomenon, first discussed in [6].

As is common, we generate random graphs and in particular random power law graphs, according to expected degree sequences [4].

Definition 2: ([4]) Let $w = (w_1, w_2, \ldots, w_n)$ be an expected degree sequence satisfying $\max\{w_k^2\} \leq \sum_{1 \leq k \leq n} w_k$. We say $G = (V, E)$ is a random graph generated by the degree sequence w if edge $(i, j) \in E$ is present with probability $w_i w_j / \sum_{1 \leq k \leq n} w_k$.

Definition 3: ([4]) A random graph generated by Definition 2 is a random power law graph with parameter β, average

degree $\bar{d}$ and maximum expected degree M if w_i is chosen by

$$w_i = c(i_o + i)^{-1/(\beta-1)}, \tag{3}$$

where $c = \frac{\beta-2}{\beta-1}\bar{d}n^{1/(\beta-1)}$ and $i_0 = n\left(\frac{\bar{d}(\beta-2)}{M(\beta-1)}\right)^{\beta-1}$.

The well-known Erdös-Rényi graph, denoted by $G(n,p)$, is the graph where each edge is present with probability p. It is thus a random graph with expected degree sequence $w = (np, np, \ldots, np)$.

For convenience, we further introduce the following notation. Given a subset $S \subseteq V$, let the volume of S be $\mathrm{vol}(S) = \sum_{i \in S} w_i$, *i.e.*, the sum of weights of nodes in S. Similarly, define $\mathrm{vol}_k(S) = \sum_{i \in S} w_i^k$ and $\tilde{d} = \mathrm{vol}_2(G)/\mathrm{vol}(G)$.

C. Assumption on file length

The transmission time consists of two parts: propagation delay and file receiving time. The propagation delay is the time required to receive the first bit since the start of the transmission. The file receiving time is the time required to finish the transmission since then. For simplicity, we assume the file length F is large, and we ignore the propagation delay in the analysis. We note in passing that we can formally incorporate both propagation delay as well as the file receiving time in our analysis by scaling F such that the propagation delay terms will be sub-dominant to the file receiving time.

III. Algorithm and main results

We are now ready to present our algorithm and state our main results. At some initial time, the file generator (*the source*) creates the file, and advertises it on her social network. At any given time, a node either has the file (*active node*), knows about the file and wants it because one of its social-network neighbors has it (*eager node*), or is oblivious to its existence (*inactive node*).

The algorithm proceeds in three phases. In the Requesting Phase, eager nodes use their social network to request the file from active nodes – if knowledge of geographic location is available, nodes favor (geographically) nearby active nodes. In the Scheduling Phase, again the social network is used to schedule a sequence of transmissions whereby each eager node is assigned a transmission node from which it will obtain the file. In the Transmission Phase, nodes transmit the file to their appointed requestors, employing established routing techniques [2]. This final third phase is conceptually distinct from the first two phases, and it is important to emphasize this point here. The routing techniques used are independent of the social network structure, and follow the multi-hop ad hoc network protocols described in, e.g., [1], [2]. Thus, while the requesting and scheduling in Phases 1 and 2 are constrained by the social network, the routing in Phase 3 is not.

We present a single algorithm that accommodates two settings: in the first, simpler setting, nodes have no notion of geography, and may not request the file from active nodes more than a single hop away on their social network. In the second setting, nodes are aware of geography and hence distance, and "prefer" to request the file from geographically

nearby nodes. Moreover, they are allowed to search for such nearby nodes beyond their immediate neighbors in the social network.

Our algorithm accommodates both settings – the first, by adjusting the "preferred distance" to infinite and the number of search-hops to 1, and the second, by limiting the preferred distance, and by expanding the number of allowed search-hops. In Section III-B we consider the first setting: no geographic information available. We show that for most social networks, our algorithm gives $\sqrt{n}$-scaling. We consider the second setting in Section III-C, where nodes have access to geographic position information. We show that again for many social networks, the dissemination time can be further reduced to scale more slowly than $\sqrt{n}$.

A. Algorithm

Our algorithm takes the input as the diameter of the social network, D, as well as two parameters which we specify: ϵ, and $\mathcal{L}$, whose roles are as follows. Nodes are allowed to search for another node in the social network from which to download the file, at a distance of at most $2\epsilon D + 1$ hops away. Thus if $\epsilon = 0$, they cannot look beyond a single hop away, and if $\epsilon = 0.5$, they have access to the entire social network. Thus the parameter ϵ controls the search depth. The parameter $\mathcal{L}$ is used to exploit geographic proximity: most nodes will download the file from nodes that are at a geographic distance of at most $\mathcal{L}$. If nodes have no notion of geography, we set $\mathcal{L} = \infty$, hence all nodes are within $\mathcal{L}$. Otherwise, we set $\mathcal{L}$ to a smaller value.

Given parameters $(\epsilon, \mathcal{L}, D)$ as described above, the algorithm finds active nodes from which eager nodes can download the file. This is accomplished through coordination through the social network.

The main idea is the following: eager nodes send requests to one of their social-network neighbors with the file. Since a single node may get many such requests, it does not serve all of them, but rather finds other active nodes nearby in the social network to serve them, and also enlists the receiving nodes themselves to forward along the file. The theorems given in Sections III-B and III-C show that for the specific choices of parameters ϵ and $\mathcal{L}$ given, the algorithm succeeds in delivering the file to all nodes, and moreover does so in the advertised time scaling.

When $\mathcal{L}$ is set to a non-infinite value, it may not always be possible for nodes to obtain the file from geographically proximate neighbors – for instance, suppose the generator has no neighbors in her geographic proximity. In such cases, we allow file transfers that exceed geographic distance $\mathcal{L}$, and these happen from two or one-hop neighbors on the social network. We call transfers within geographic distance $\mathcal{L}$, $\mathcal{L}$-transfers, and all other transfers S-transfers, since they are near in the social-network distance. Similarly we refer to $\mathcal{L}$-requests and S-requests.

ALGORITHM 1:

Input: parameter ϵ, distance threshold $\mathcal{L}$, and the diameter of the social network D.

Requesting Phase: Consider an eager node, x, at time t.

Step 1: Let $\mathcal{N}_x(t)$ denote node x's $2\epsilon D + 1$-neighborhood in the social-graph at time t. Let $\mathcal{N}_x^{\mathcal{L}}(t) \subseteq \mathcal{N}_x(t)$ be the set of nodes in $\mathcal{N}_x(t)$ that have the file and whose Euclidean (geographic) distance to x does not exceed $\mathcal{L}$.

Step 2: If $\mathcal{N}_x^{\mathcal{L}}(t)$ is not empty, x sends an $\mathcal{L}$-request to a randomly picked node in $\mathcal{N}_x^{\mathcal{L}}(t)$.

Step 3: If $\mathcal{N}_x^{\mathcal{L}}(t)$ is empty and the distance from x to the source on the social-graph is smaller than $\epsilon D + 1$, then x sends an S-request to a one-hop neighbor in the social-graph which has the file.

Step 4: Otherwise, x waits and goes back to step 1 at time $t + 1$.

Scheduling Phase: Consider an active node y. It maintains two balanced binary trees, an $\mathcal{L}$-tree and an S-tree, constructed from its $\mathcal{L}$-requests and S-requests, respectively. It builds these trees by adding requesting nodes sequentially, as the requests arrive. This sequential building of the binary trees is depicted in Figure 1.

When node y receives an $\mathcal{L}$-request, node y adds the eager node to the $\mathcal{L}$-tree and asks its parent on the tree to deliver the file, and similarly for S-requests.

Transmission Phase: An eager node waits until the node designated as its transmitting node in the Scheduling Phase has the file. It then sets up a wireless transmission, and routes data through a highway system described in [2]. Note that the transmitter will have to serve at most 6 nodes: 2 from its own $\mathcal{L}$-tree, 2 from its own S-tree, and 2 from the tree it joins when it is an eager node (which could be either an $\mathcal{L}$-tree or an S-tree). Thus, we divide a time slot into six and each transmitter serves all nodes in a round robin fashion.
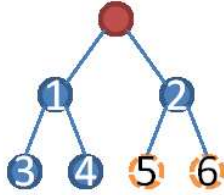

Fig. 1. Each active node maintains balanced binary trees and adds requesting nodes to trees sequentially. Suppose the active node depicted at the root gets four requests at time t, and two more at time $t + 1$. The resulting tree might look as depicted. The original active node would then serve nodes 1 and 2, subsequently node 1 would serve nodes 3 and 4, and node 2 would serve nodes 5 and 6.

B. Main results: load balancing

In this section we show that the load-balancing accomplished by the $\mathcal{L}$-binary trees is enough to give $\sqrt{n}$-scaling, without any geographic information. We show that our result holds, as long as the social network has the properties of a random power law graph with $\beta > 2$, minimum expected degree $m > 3$ and maximum expected degree M satisfying $\log(n) \ll M \ll \sqrt{n}$ (many social networks have values of β large than this – see, for example, collaboration graphs in [9]). In this case, the diameter of the social-graph is $\mathcal{O}(\log(n))$ and the size of the largest component is of $\Theta(n)$ [4][5].

As discussed, we set $\mathcal{L} = \infty$, and $\epsilon = 0$, thus nodes are only allowed to request the file from nodes at most one hop away on the social network, and they entirely ignore geography.

In this case, for any eager node x, we have $\mathcal{N}_x^{\mathcal{L}}(t) \neq \emptyset$ at the time t node x becomes eager, and hence the Requesting Phase of the algorithm uses only Step 1 and Step 2. There are only $\mathcal{L}$-requests, and thus the algorithm requires each node to transmit to at most 4 other nodes. Indeed, the point of this algorithm is to distribute the load evenly on the wireless network.

In Theorem 4, we show that the file dissemination time scales like $\sqrt{n}$ (sublinearly). In addition, we show that the performance only differs from *algorithm independent* lower bounds with a factor n^ξ for any $\xi > 0$. Since the proofs of the following two theorems are similar to those for Theorem 7 and Theorem 8, we defer the full details to the Appendix B.

Theorem 4: Consider the file dissemination problem with wireless network and social network as defined above. Suppose the file length is F. Then the file dissemination time for Algorithm 1 is

$$\mathcal{O}(\sqrt{n} \log^2(n) F) \tag{4}$$

with high probability.

Theorem 5: Consider the file dissemination problem with wireless network and social network as defined above. Suppose the file length is F. Then, for any algorithm that allows nodes to download the file from their 1- and 2-hop neighbors on the social network, the file dissemination time is lower bounded by

$$\Omega(n^{1/2 - \xi} F), \tag{5}$$

for any $\xi > 0$ with high probability.

Remark 6: Significantly, the only properties of power law graphs we use are the size of the diameter and the maximum degree. Specifically, given a graph G with diameter ℓ_{max} and maximum degree d_{max}, the file dissemination time is $\mathcal{O}(\sqrt{n} \log(d_{max}) \ell_{max} F)$ if nodes are only allowed to download the file from nodes at most 2 hops away. The proof of this follows immediately from the proof of the theorem.

C. Main results: exploiting geography

Intuitively, increasing the number of geographically proximal downloads should decrease transmission time. We show that this can be accomplished, at the cost of deeper searching of the social network, as long as the social network has the properties of a random power law graph with $\beta > 3$ (again, many graphs have this property, see, e.g., the collaboration graphs in [10]). We assume that the minimum expected degree is $m = K \log(n)$ where K is a constant greater than 10, and the maximum expected degree is M, satisfying $\log^2(n) \ll M \ll \sqrt{n}$. Thus, almost all nodes are in the largest component and the diameter of the graph is $D \approx \log_{\tilde{d}}(n)$ [4][5] (recall the definition of $\tilde{d}$ from Section II).

Setting ϵ to a positive value translates to allowing nodes to search for an active node in their $2\epsilon \log_{\tilde{d}}(n) + 1$-neighborhood, and because of our load-balancing architecture, ultimately

download the file from nodes in their $4\epsilon \log_{\tilde{d}}(n) + 2$ neighborhood. With more active nodes available, eager nodes can more easily find geographically proximal active nodes. We set $\mathcal{L} = 8\sqrt{n^{1-\epsilon'} \log(n)/\sigma\pi}$ for any $\epsilon' < \epsilon$. The value of ϵ is chosen to be small, $\epsilon < 1/10$, allowing nodes to search a neighborhood that is large, but nevertheless a vanishing fraction of the size of the entire network.

As load-balancing alone was able to achieve file dissemination time scaling of $\sqrt{n}$, we show now that by additionally exploiting geography, the file dissemination time can be further reduced by a factor $n^{\epsilon/2}$ compared to the result in Theorem 4. Proofs of the two theorems can be found in Section V.

Theorem 7: Suppose the source is chosen uniformly at random from the nodes in the largest component and the file length is F. Consider the setting described above. Then the file dissemination time under Algorithm 1 with parameter $0 < \epsilon < 0.1$ is

$$\mathcal{O}(\sqrt{n^{1-\epsilon'}} \log^{2.5}(n)F), \tag{6}$$

for any $\epsilon' < \epsilon$ with high probability.

Theorem 8: Consider the file dissemination problem under the setting described above. Let F be the file length. Then, for any algorithm that allows nodes to download the file from their $4\epsilon \log_{\tilde{d}}(n) + 2$-neighborhood with $\epsilon < 0.1$, the file dissemination time is lower bounded by

$$\Omega(n^{1/2-2\epsilon-\xi}F), \tag{7}$$

with high probability for any $\xi > 0$.

IV. Random Placement and Random Graphs

In preparation for the proof in the next section, we give some lemmas that characterize the behavior of randomly placed nodes in a square, and also give properties of random graphs.

A. Results about random placement of nodes in a square

Two properties in particular, are important. For our scheme to work, we need to show that with overwhelming probability, we will not have a very high clustering of nodes (some clustering will occur). We also need to show that when nodes look in their social network for geographically proximate active nodes, they will be able to find at least one, with high probability. The next two lemmas show precisely these properties.

In the first lemma, we show we control the minimum distance between a node and k other nodes. We use this lemma to ensure that each node can find a node close to it on the wireless-square. In the second lemma, we show a concentration result about the number of nodes falling into a small rectangle, thus showing it is not too big. The proofs of these lemmas are also available in Appendix B.

Lemma 9: Place $k + 1$ nodes on a square of width $\sqrt{n}$ independently and uniformly. Let τ be the minimum distance from the first node to the others. Then, we have

$$\mathbb{P}(\tau \geq \sqrt{64n\log(n)/\pi k}) \leq n^{-2}. \tag{8}$$

Lemma 10: Place n nodes on a square of width $\sqrt{n}$ independent and uniformly. Given a rectangle of area A where $A = \omega(\log(n))$, let X be the number of nodes in the rectangle. Then,

$$\mathbb{P}(X \geq 2A) \leq n^{-2}. \tag{9}$$

B. Results about the neighborhood behavior of random graphs

In the following lemma, we address the relation between weights and the number of neighbors. Specifically, we show that if a node has weight w_i greater than $10 \log(n)$, then the number of one-hop neighbors the node can reach in the social-graph is between $w_i/2$ and $2w_i$. We use this lemma as it provides a relationship between weights and the number of nodes.

Lemma 11: Suppose $w_i \geq 10 \log(n)$. Let X be the number of one-hop neighbors in the social-graph of node i. Then, $w_i/2 < X < 2w_i$ with probability $1 - o(n^{-1})$.

The next two lemmas characterize the local behavior of random power law graphs. Specifically, we are interested in how the size of neighborhoods of nodes in the largest component grows. We show that for any node in the largest component, the number of nodes in a small neighborhood grows like a factor $\tilde{d}$ if we explore one more step. We prove this by providing upper and lower bounds that only differ by a factor of n^ξ for any $\xi > 0$. The proofs are shown in Appendix A.

Lemma 12: Consider a random power law graph with parameter $\beta > 3$. Suppose the minimum expected degree is $m = K \log(n)$ for some $K \geq 10$ and the maximum expected degree is $M \gg \log^2(n)$. Then, there are at least $\sigma n^{\epsilon'}$ nodes in a node's $\epsilon \log_{\tilde{d}}(n)$-neighborhood with probability $1 - o(n^{-1})$, for any $\epsilon' < \epsilon < 0.1$. Here, σ is a constant depending on β and K.

Lemma 13: Consider a random power law graph with parameter $\beta > 3$. Suppose the minimum expected degree is $m = K \log(n)$ for some $K \geq 10$, the maximum expected degree is $M \gg \log^2(n)$, and $\epsilon < 0.4$. Consider a node either picked randomly or with weight smaller than W. Then, there are at most $2W\tilde{d}^\lambda n^{\epsilon'}/\log(n)$ nodes in this node's $\epsilon \log_{\tilde{d}}(n) + \lambda$-neighborhood, with probability $1 - \mathcal{O}(\log^{-1}(n))$, for any $\epsilon' > \epsilon$ and any fixed constant λ where

$$W = \begin{cases} \log^{\beta/\beta-3}(n) & \text{if } 3 < \beta \leq 4 \\ \max\{\log^{5/\beta-4}(n), \log^2(n)\} & \text{if } 4 < \beta \end{cases} \tag{10}$$

V. Performance Analysis

A. Proof of Theorem 7

In this section, we first prove Theorem 7 which states the performance of our algorithm when geographic information is available, and when nodes can download the file from a neighborhood of radius $4\epsilon \log_{\tilde{d}}(n) + 2$. The proof of the more simple load-balancing case (where we set $\mathcal{L} = \infty$ and $\epsilon = 0$) is essentially a consequence of this proof – for the full details we refer to Appendix B. Specifically, we show the file dissemination time is roughly $\sqrt{n^{1-\epsilon}}F$.

The proof of the theorem consists of two main parts: showing the existence of a geographically nearby neighbor in the wireless-graph and the analysis of transmission rates. In addition, the transmission phase of our algorithm relies on some routing results from [2] and [12] which we summarize here. For the full details, we refer readers to those individual papers. In the routing scheme, packets are routed through a highway system consisting of horizontal highways and vertical highways. Each highway serves nodes in a stripe on the wireless-square. An illustration is shown in Fig. 2. The results in [2] and [12] guarantee the following properties of this highway system.

1) Nodes can reach their highways in a hop of length $\mathcal{O}(\log(n))$.
2) The highways are almost straight. For example, if a flow on a horizontal highway starts from x-coordinate a_1 with destination at x-coordination $a_2 > a_1$, it will not reach any node with x-coordinate smaller than $a_1 - H$ where $H = \mathcal{O}(\log(n))$.
3) Highway nodes can communicate with neighboring highway nodes with a constant rate. A highway node serves flows through it with equal rate.
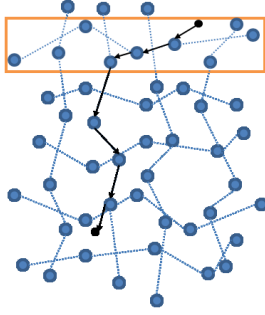

Fig. 2. An illustration of the highway system and routing. Packets are first routed through horizontal highways to vertical highways corresponding to destinations.

We now move to the proof of the theorem. We first state the existence of an "intermediate node" in the following lemma.

Lemma 14: Consider a random power law graph with $\beta > 3$ and $m = K \log(n)$ where $K > 10$ is a constant. For some $\epsilon < 0.1$, consider a node x in the largest component, such that the distance from x to the source on the social network is greater than $\epsilon \log_{\bar{d}}(n)$. Then, there exists a node y which satisfies the follows with probability at least $1 - o(n^{-1})$:

1) y is in the $2\epsilon \log_{\bar{d}}(n)+1$-neighborhood of x in the social-graph.
2) The distance from y to the source on the social network is smaller than that from x to the source.
3) The Euclidean distance from x to y on the wireless-square is smaller than $\mathcal{L}$.

Proof. We first show that there exist $\sigma n^{\epsilon'}$ nodes satisfying 1) and 2) with probability $1-o(n^{-1})$. Let d_x be the distance from x to the source on the social network. Since $d_x > \epsilon \log_{\bar{d}}(n)$, there exists a node z such that the distance from z to x on the graph is $\epsilon \log_{\bar{d}}(n) + 1$ and the distance from z to the source

on the graph is $d_x - \epsilon \log_{\bar{d}}(n) - 1$. Therefore, nodes in the $\epsilon \log_{\bar{d}}(n)$-neighborhood of z in the social-graph satisfy 1) and 2). In addition, by Lemma 12, the size of such a neighborhood is greater than $\sigma n^{\epsilon'}$ with probability $1 - o(n^{-1})$.

Thus, by Lemma 9, there exists a node y among the $\sigma n^{\epsilon'}$ nodes whose Euclidean distance to x on the wireless-square is smaller than $\mathcal{L}$ with probability $1 - o(n^{-1})$. ∎

Proof. (Theorem 7) Recall that our algorithm classifies transmissions as those chosen because they are geographically within distance $\mathcal{L}$, called $\mathcal{L}$-transmissions, and those chosen because they are within two hops on the social network, called S-transmissions. S-transmissions are those whose Euclidean distances between transmitters and receivers on the wireless-square are not guaranteed to be less than $2\mathcal{L}$, as are $\mathcal{L}$-transmissions. Note that the number of S-transmissions is smaller than the number of nodes in $\epsilon \log_{\bar{d}}(n)$-neighborhood of the source in the social-graph which is smaller than $2W n^{\epsilon''}/\log(n)$ for any $\epsilon'' > \epsilon$ with high probability by Lemma 13.

Now, we bound the number of flows through a highway node at any time. Consider a transmission between two nodes with Euclidean distance less than $2\mathcal{L}$. By the fact that highways are almost straight and the first and last hops are of length $\mathcal{O}(\log(n))$, the transmission passes through a horizontal (vertical) highway node only if the horizontal (vertical) distance between the transmitter (receiver) and the node is smaller than $3\mathcal{L}$ on the wireless-square. In other words, $\mathcal{L}$-transmissions through a horizontal (vertical) highway node must fall in a rectangle of side $6\mathcal{L} \times h$ in the corresponding horizontal (vertical) strip where h is a constant provided in [2]. Since, by Lemma 10, the total number of nodes falling into this region is $\mathcal{O}(\mathcal{L})$ with probability $1-o(n^{-1})$ and each node generates at most a constant number of flows, using the union bound we can conclude that all highway nodes have at most $\mathcal{O}(\mathcal{L})$ $\mathcal{L}$-flows with probability $1 - o(1)$. In addition, since there are at most $2W n^{\epsilon''}/\log(n)$ S-transmissions, the total number of flows through each highway node is $\mathcal{O}(\mathcal{L})$ with probability $1 - o(1)$. Therefore, each flow has a rate $\Omega(1/\mathcal{L})$ with high probability and each node can receive the file in $c_0 \mathcal{L} F$ time slots for some constant $c_0 > 0$ from the time when the transmission begins.

We prove the theorem by induction on k: the distance from a node to the source on the social-graph. Let $\mathcal{N}_k$ denote nodes whose distance to the source is k on the social-graph. The claim of the induction is that a node in $\mathcal{N}_k$ can receive the file in at most $kc_0 \log_2(n)\mathcal{L}F$ time slots. By our notation, $\mathcal{N}_0$ is the source node. First note, that the base case $k = 1$ of the induction clearly holds. Now, we suppose it is true for $k - 1$ and consider nodes in $\mathcal{N}_k$. Note that no nodes in $\mathcal{N}_k$ are inactive at time $(k - 1)c_0 \log_2(n)\mathcal{L}F$. Further, by Algorithm 1 and Lemma 14, all nodes in $\mathcal{N}_k$ can request the file, according to the algorithm, from an active node in $\cup_{i=0}^{k-1}\mathcal{N}_i$. Thus, these nodes have to wait at most $\log_2(n) - 1$ successful transmissions before starting to receive the file, since the depth of any binary tree is at

most $\log_2(n)$. Therefore, they can receive the file before time $kc_0 \log_2(n)\mathcal{L}F$. Hence, by induction, the file dissemination time is $\mathcal{O}(\sqrt{n^{1-\epsilon'}}\log^{2.5}(n)F)$ as the diameter of the social-graph is $\mathcal{O}(\log(n))$. ∎

B. Proof of Theorem 8

We proceed by first providing some definitions and a lemma. Given a transmission pair with rate r over an Euclidean distance ρ on the wireless-square, define the bit-meter rate of the transmission pair as $r\rho$. The total bit-meter product a network can transmit is the supremum of the sum of bit-meter products of all transmission pairs.

Lemma 15: The total bit-meter product the network can transmit in a time slot is $\Theta(n)$.

Proof. From (1), we know the bit-meter product a transmission pair (x_i, x_j) can transmit is

$$||x_i - x_j|| \log\left(1 + \frac{P\ell(x_i, x_j)}{N_0 + \sum_{k \neq i} P\ell(x_k, x_j)}\right)$$
$$\leq P\ell(x_i, x_j)||x_i - x_j||/N_0. \quad (11)$$

Recall that $\ell(x_i, x_j)||x_i - x_j||$ is bounded by a constant either for $\gamma > 0$ or $\gamma = 0$ and $\alpha > 2$. Since there are at most $n/2$ transmission pairs, the total bit-meter product the system can transmit is $\Theta(n)$ in a time slot. ∎

To prove the lower bound, we place no restrictions on computation or communication overhead. Moreover, we make (overly) optimistic assumptions throughout in order to guarantee a bound. For instance, we assume nodes only download from their nearest social-network neighbors.

Proof. (Theorem 8) Define the transport load as the infimum of the total bit-meter product required to disseminate the file under the problem setting. To apply Lemma 15, we just need to show that the transport load is $\Omega(n^{3/2-2\epsilon-\xi}F)$ with probability $1 - o(1)$.

Let $\mathcal{M}$ be the set of nodes in the largest component with expected degree in the range $[K\log(n)\ 2K\log(n)]$. Then,

$$|\mathcal{M}| \approx \frac{\int_{K\log(n)}^{2K\log(n)} x^{-\beta} n\, dx}{\int_{K\log(n)}^{M} x^{-\beta} dx} = (1 + o(1))(1 - 2^{1-\beta})n. \quad (12)$$

since almost all nodes are in the largest component.

Fix any $\epsilon' > \epsilon$. Let $\mathcal{N}_i$ be the set of nodes that node i can reach in $4\epsilon \log_{\tilde{d}}(n) + 2$ hops in the social-graph. Let X_i be the indicator that the Euclidean distance from node i to $\mathcal{N}_i$ on the wireless-square is smaller than $\sqrt{n/2\pi W \tilde{d}^2 n^{4\epsilon'}}$. Therefore, we have, for $i \in \mathcal{M}$ and n large enough and some constant c_1,

$$\mathbb{P}(X_i = 1) \leq \mathbb{P}(\{X_i = 1\} \cap \{|\mathcal{N}_i| \leq 2W\tilde{d}^2 n^{4\epsilon'}/\log(n)\})$$
$$+ \mathbb{P}(|\mathcal{N}_i| > 2W\tilde{d}^2 n^{4\epsilon'}/\log(n))$$
$$\leq 1/\log(n) + \mathcal{O}(1/\log(n)) \leq c_1/\log(n)$$

since the probability that a node is close is smaller than $1/2W\tilde{d}^2 n^{4\epsilon'}$ and the second term comes from the probability that $|\mathcal{N}_i| \geq 2W\tilde{d}^2 n^{4\epsilon'}/\log(n)$. Therefore, we have

$$\mathbb{E}\left[\sum_{i \in \mathcal{M}} X_i\right] \leq 2c_1(1 - 2^{1-\beta})n/\log(n) \quad (13)$$

We claim that $\mathbb{P}(|\sum_{i \in \mathcal{M}} X_i - \mathbb{E}[\sum_{i \in \mathcal{M}} X_i]| \geq n/\log^{1/3}(n)) = o(1)$. Indeed, by Chebyshev's inequality, we have

$$\mathbb{P}\left(\left|\sum_{i \in \mathcal{M}} X_i - \mathbb{E}\left[\sum_{i \in \mathcal{M}} X_i\right]\right| \geq n/\log^{1/3}(n)\right)$$
$$\leq \frac{\mathbb{E}[(\sum_{i \in \mathcal{M}} X_i)^2]}{n^2/\log^{2/3}(n)}$$
$$\leq \frac{c_1 n^2/\log(n)}{n^2/\log^{2/3}(n)} = o(1) \quad (14)$$

where the last inequality follows from $\mathbb{E}[X_i X_j] \leq \mathbb{E}[X_i]$.

By the above claims, $|\mathcal{M}| = \Theta(n)$ while the number of nodes with geographically close neighbors in the wireless-square is $o(n)$. Hence, the transport load is $\Omega(n^{3/2-2\epsilon-\xi}F)$. ∎

VI. CONCLUSIONS

New technology (smartphones, etc.) has made content creation easy – just a press of a button. Social networks, meanwhile, make wide dissemination of the *knowledge of* that file, just as easy – a press of another button. Yet actual dissemination of large files to many users can seriously burden a wireless network. In the WAN setting, the time to disseminate must scale linearly in the number of users. In this paper, we consider simple, low-overhead file dissemination algorithm that exploits peer-to-peer capabilities of many smartphones and similar devices, and, critically, exploits the very social networks that spread knowledge of the file. We give a load-balancing algorithm that uses the social network to schedule transmissions so that spatial-capacity of the ad hoc network is exploited without creating congestion or hot spots. We show that dissemination time scales like $\sqrt{n}$ — significantly slower than the linear time for WAN. Then, we show that if nodes have knowledge of geographic position, this can be exploited to further decrease file dissemination time. Finally, we show in both cases that our algorithm performs close to an algorithm-independent lower bound.

VII. APPENDIX A

A. Proof for Lemma 12

We first quote lemma 3.2 from [4]. This useful lemma addresses how a neighborhood of a set in the random power law graph grows. Specifically, if we have two sets S and T, what is the sum of weights of neighbors of S which are also in T? One important application is the setting where $T \approx G$, i.e., T is almost the entire graph. In this case, we get an increase factor of roughly $\tilde{d}$.

Lemma 16: ([4]) Given a random graph and two subsets S and T, if

$$\frac{2c}{\delta^2}\frac{\mathrm{vol}_3(T)}{\mathrm{vol}_2^2(T)} \leq \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)}, \tag{15}$$

$$\frac{\mathrm{vol}(S)}{\mathrm{vol}(G)} \leq \delta\frac{\mathrm{vol}_2(T)}{\mathrm{vol}_3(T)}, \tag{16}$$

we have

$$\mathrm{vol}(\Gamma(S) \cap T) \geq (1 - 2\delta)\frac{\mathrm{vol}_2(T)}{\mathrm{vol}(G)}\mathrm{vol}(S) \tag{17}$$

with probability $1 - e^{-c}$ where $\Gamma(S)$ is the set of one-hop neighbors of S.

Using this lemma, we provide a proof to Lemma 12, which we used to lower bound the size of a node's immediate neighborhood.

Proof. (Lemma 12) Consider node x's neighborhood. Let S_i be the set of nodes whose distance to x is i and $S_0 = \{x\}$. We will show that $\mathrm{vol}(S_{i+1}) \geq (1 - 2\delta)\tilde{d}\mathrm{vol}(S_i)$ for $\delta = 1/4$ with probability $1 - o(n^{-2})$. To do this we need to apply Lemma 16 inductively and choose $c = 3\log(n)$. We may assume $\mathrm{vol}(S_i) < n^{2\epsilon}$ in the first $\epsilon\log_{\tilde{d}}(n)$ steps. Since $\frac{\mathrm{vol}_2(T)}{\mathrm{vol}_3(T)}\mathrm{vol}(G) = \Omega(n/M)$ for all T, (16) holds for all $\beta > 3$. We have only to verify (15).

First notice that $\mathrm{vol}(S_1) = \Omega(\log^2(n))$ with probability $1 - o(n^{-1})$. This is true since, by Lemma 11, the node x has at least $K\log(n)/2$ neighbors with probability $1 - o(n^{-1})$ and each neighbor has weight at least $K\log(n)$.

We next verify (15) for $\beta > 4$. Let T_i be the set of all potential nodes whose distance to x is $i + 1$, i.e., $T_i = G \setminus \cup_{k=0}^i S_k$. Then, $\mathrm{vol}_2(T_i) = (1 + o(1))\mathrm{vol}_2(G)$ and $\mathrm{vol}_3(T_i) = (1 + o(1))\mathrm{vol}_3(G)$. Thus, $\frac{\mathrm{vol}_3(T_i)\mathrm{vol}(G)}{\mathrm{vol}_2^2(T_i)} = \Theta(1)$. Therefore, we have the result by induction.

For the case $3 < \beta \leq 4$, let $T_i^{(k)}$ be the intersection of T_i and the set of nodes with weight smaller than $k\log(n)$. Then, we have

$$\frac{2c}{\delta^2}\frac{\mathrm{vol}_3(T_i^{(k)})}{\mathrm{vol}_2^2(T_i^{(k)})}\mathrm{vol}(G) = \mathcal{O}(\log(n)) \tag{18}$$

and

$$\frac{\mathrm{vol}_2(T_i^{(k)})}{\mathrm{vol}_2(G)} \approx \frac{\int_{K\log(n)}^{k\log(n)} x^{2-\beta}dx}{\int_{K\log(n)}^M x^{2-\beta}dx}$$
$$= (1 + o(1))(1 - (\frac{k}{K})^{3-\beta}). \tag{19}$$

Therefore, by (18), we have (16) is true by induction. On the other hand, $\mathrm{vol}_2(T_i^{(k)})/\mathrm{vol}(G) \approx \tilde{d}$ as k becomes large enough, by (19).

With the above results, we can conclude that the size of the neighborhood grows by roughly a factor of $\tilde{d}/2$ with probability $1 - o(n^{-2})$ for each step, from the second one to the $\epsilon\log_{\tilde{d}}(n)^{th}$ step. Since $\mathrm{vol}(S_{\epsilon\log_{\tilde{d}}(n)-1}) \geq 2\sigma n^{\epsilon(1-o(1))}$ for some constant σ, there are at least $\sigma n^{\epsilon'}$ nodes within distance $\epsilon\log_{\tilde{d}}(n)$ of x in the social network. ∎

B. Proof for Lemma 13

The proof of Lemma 13 depends on the following lemma from [5], that provides large deviation results for both an upper bound and a lower bound for the sum of Bernoulli random variables.

Lemma 17: ([5]) Let X_i be a Bernoulli random variable with parameter p_i. Suppose $\{X_i\}$ are independent. Let $X = \sum_{i=1}^n a_iX_i$ and $\nu = \sum_{i=1}^n a_i^2p_i$. Then, we have

$$\mathbb{P}(X \leq \mathbb{E}[X] - c) \leq \exp(-c^2/2\nu) \tag{20}$$

$$\mathbb{P}(X \geq \mathbb{E}[X] + c) \leq \exp(-c^2/2(\nu + ac/3)) \tag{21}$$

where $a = \max\{a_1, a_2, \ldots, a_n\}$.

Proof. (Lemma 13) We first state the flow of the proof. In the beginning, we show that we only need to consider an initial node x with weight W. We next define S_i as the set of nodes at distance i from node x and show that, for any $\delta > 0$,

$$\mathrm{vol}(S_i) \leq W((1 + \delta)\tilde{d})^i. \tag{22}$$

We in fact show $\mathrm{vol}(S_{i+1}) \leq (1+\delta)\tilde{d}\mathrm{vol}(S_i)$ with probability $1 - \mathcal{O}(\log^{-2}(n))$. To do so, we construct a set T^i which contains nodes with large weight and show $T^i \cap S_{i+1} = \emptyset$ with overwhelming probability. On the other hand, we use Lemma 17 to bound the sum of weights in S_{i+1} contributed by nodes with small weight. To do this, we have to consider three cases depending on β.

We now present the details of the proof. We first show the condition, $\mathrm{vol}(S_0) \leq W$. Since $S_0 = \{x\}$, we need to show the weight of x does not exceed W if x is picked randomly.

$$\int_{\log^2(n)}^M x^{-\beta}dx/C_1 = \frac{1 + o(1)}{C_1(\beta - 1)}(\log^2(n))^{1-\beta}$$
$$= o(\log^{-2}(n)), \tag{23}$$

where C_1 is the normalization constant $\int_m^M x^{-\beta}dx$. Thus, a randomly picked node x has weight smaller than W with high probability. Since by standard coupling arguments we see that the growth of the neighborhood of x is dominated by a node with weight W, we simply take the weight of x to be W in what follows.

We turn to show $\mathrm{vol}(S_{i+1}) \leq (1 + \delta)\tilde{d}\mathrm{vol}(S_i)$ with probability $1 - \mathcal{O}(\log^{-2}(n))$. First, we give some definitions. Let $\tilde{m}_i = (\mathrm{vol}(S_i)\log^\beta(n))^{1/(\beta-2)}$ and define T^i to be the set of nodes with weight greater than $\tilde{m}_i$. We first show that $T^i \cap S_{i+1} = \emptyset$ with probability $1 - \mathcal{O}(\log^{-2}(n))$. Indeed,

$$\mathbb{P}(S_{i+1} \cap T^i \neq \emptyset) \leq \mathrm{vol}(S_i)\mathrm{vol}(T^i)/\mathrm{vol}(G)$$
$$\approx \mathrm{vol}(S_i)\int_{\tilde{m}_i}^M x^{1-\beta}ndx/C_1\mathrm{vol}(G)$$
$$= \mathcal{O}(\log^{-2}(n)), \tag{24}$$

where the first inequality follows from the union bound. Therefore, with high probability, $S_{i+1} \cap T^i = \emptyset$.

Define $\tilde{T}_i$ to be the set of unexplored nodes with weight smaller than $\tilde{m}$ in the i-th step, i.e., $\tilde{T}_i = V \setminus (T^i \cup \cup_{j=0}^i S_j)$. Thus, S_{i+1} is a subset of $\tilde{T}_i$. For simplicity, we consider a

larger set T_i which includes $\tilde{T}_i$ and virtual nodes $\tilde{V}_i$ where $\tilde{V}_i$ is chosen such that the number of nodes and their weights are the same as those in $\cup_{j=0}^i S_j$. We allow nodes in S_i to connect to nodes in $\tilde{V}_i$ and, therefore, have a looser upperbound on $\mathrm{vol}(S_{i+1})$. We first give some properties of T_i. Specifically, we claim $\mathrm{vol}(T_i) = (1 + o(1))\mathrm{vol}(G)$ and $\mathrm{vol}_2(T_i) = (1 + o(1))\mathrm{vol}_2(G)$. Indeed,

$$
\begin{aligned}
\mathrm{vol}(T_i) &= \mathrm{vol}(G) - \mathrm{vol}(T^i) \\
&\geq \mathrm{vol}(G) - \int_{\log^{\beta/(\beta-2)}(n)}^{M} x^{1-\beta} n dx / C_1 \\
&= \mathrm{vol}(G) - \mathcal{O}(n \log^{-1}(n)) \\
&= (1 + o(1))\mathrm{vol}(G),
\end{aligned}
\tag{25}
$$

where the inequality follows from $\mathrm{vol}(S_i)$ will increase by a factor at least $\tilde{d}/2$ in each step as shown in Lemma 12. Similarly, we have $\mathrm{vol}_2(T_i) = (1 + o(1))\mathrm{vol}_2(G)$.

Next, we give some properties of S_{i+1}. Our goal is to find the expected weight of S_{i+1}, $\mathbb{E}[Y_i]$, and the variable ν_i (defined below) to apply Lemma 17. To do this, define X_j as the indicator function that node j is in S_{i+1}. Thus, by union bound and a fact (in the proof of Lemma 3.2 in [4]), we have

$$
\begin{aligned}
\mathrm{vol}(S_i)w_j/\mathrm{vol}(G) - (\mathrm{vol}(S_i)w_j/\mathrm{vol}(G))^2 \\
\leq \mathbb{P}(X_j = 1) \leq \mathrm{vol}(S_i)w_j/\mathrm{vol}(G).
\end{aligned}
\tag{26}
$$

Let Y_i be the volume of S_{i+1}, i.e., $Y_i = \sum_{j \in T_i} w_j X_j$. Thus, we have

$$
\mathbb{E}[Y_i] = (1 + o(1))\mathrm{vol}(S_i)\mathrm{vol}_2(T_i)/\mathrm{vol}(G).
\tag{27}
$$

Similarly, define $\nu_i = \sum_{j \in T_i} w_j^2 \mathbb{P}(X_j = 1)$. We have

$$
\nu_i = (1 + o(1))\mathrm{vol}(S_i)\mathrm{vol}_3(T_i)/\mathrm{vol}(G).
\tag{28}
$$

Using the properties of T_i and S_{i+1}, we now show the inductive step, namely: $\mathrm{vol}(S_{i+1}) \leq (1 + \delta)\tilde{d}\,\mathrm{vol}(S_i)$ with probability $1 - o(\log^{-2}(n))$. Recalling Lemma 17, we have

$$
\mathbb{P}(Y_i > \mathbb{E}[Y_i] + \kappa_i) \leq \exp\left(-\frac{\kappa_i^2}{2(\nu_i + \tilde{m}_i \kappa_i/3)}\right).
\tag{29}
$$

We need to consider three cases which are $3 < \beta < 4$, $\beta = 4$, and $\beta > 4$. In each case, we first estimate ν_i and then compare $c\tilde{m}_i$ and $\sqrt{c\nu_i}$. According to the above comparison, we specify κ_i for each case and conclude our desired result. Let $c = 10 \log\log(n)$ and consider the three cases.

Case 1 ($3 < \beta < 4$): First note that

$$
\begin{aligned}
\nu_i &\approx \mathrm{vol}(S_i) \int_m^{\tilde{m}_i} x^{3-\beta} n dx / C_1 \mathrm{vol}(G) \\
&= \mathcal{O}(\mathrm{vol}(S_i)\tilde{m}_i^{4-\beta} \log^{\beta-2}(n)).
\end{aligned}
\tag{30}
$$

Therefore, $c\tilde{m}_i \gg \sqrt{c\nu_i}$ for n sufficiently large. Hence, choosing $\kappa_i = c\tilde{m}_i$, we have

$$
\mathbb{P}(Y_i > \mathbb{E}[Y_i] + \kappa_i) \leq \exp\left(-\frac{\kappa_i^2}{4\tilde{m}_i \kappa_i}\right) = o(\log^{-2}(n)).
\tag{31}
$$

Note that by (27), we need only to show that $\kappa_i \leq \delta \mathrm{vol}(S_i)\mathrm{vol}_2(T_i)/\mathrm{vol}(G)$. This suffices to show

$$
\mathrm{vol}(S_i) \geq \left(\frac{c\mathrm{vol}(G)}{\delta \mathrm{vol}_2(T_i)}\right)^{\beta-2/\beta-3} \log^{\beta/\beta-3}(n).
\tag{32}
$$

But this is true since the initial weight is greater than $\log^{\beta/\beta-3}(n)$ and it increases by a factor of at least $\tilde{d}/2$ in each step.

Case 2 ($\beta = 4$): We have $\nu_i = \mathcal{O}(\mathrm{vol}(S_i) \log(\tilde{m}_i) \log^2(n))$ and $c\tilde{m}_i \gg \sqrt{c\nu_i}$. Hence, with similar computation as that described in Case 1, we have the desired result.

Case 3 ($\beta > 4$): We have $\nu_i = \Theta(\mathrm{vol}(S_i) \log^2(n))$. By direct computation, we have $\sqrt{c\nu_i} \gg c\tilde{m}_i$ provided the initial weight is greater than $\log^{5/\beta-4}(n)$. Hence, we choose $\kappa_i = \sqrt{c\nu_i}$. Similarly, we just need to show that $\kappa_i \leq \delta \mathrm{vol}(S_i)\mathrm{vol}_2(T_i)/\mathrm{vol}(G)$. This is true since, by (28),

$$
\mathrm{vol}(S_i) \geq \frac{c\mathrm{vol}_3(T_i)\mathrm{vol}(G)}{\delta^2 \mathrm{vol}_2(T_i)^2} = \Theta(\log\log(n)).
\tag{33}
$$

Note that the probability of failure in each step is $\mathcal{O}(\log^{-2}(n))$ and there are at most $\epsilon \log_{\tilde{d}}(n) + \lambda$ steps. Thus, the sum of weights of nodes in $\epsilon \log_{\tilde{d}}(n) + \lambda$-neighborhood of x is at most $2W\tilde{d}^\lambda n^{\epsilon(1+o(1))}$. We conclude that the desired result holds with probability $1 - \mathcal{O}(\log^{-1}(n))$ as each node has weight at least $\log(n)$. ∎

VIII. APPENDIX B

A. Proofs of lemmas in Section IV

In this section, we first show Lemma 9, Lemma 10, and Lemma 11. The proofs of these lemmas use techniques for balls and bins problems. To solve a problem like these, in general, we first find a proper target function and write the target function as a sum of indicator functions. We next use a large deviation result, e.g. Lemma 17, to show the target function is concentrated around its mean. The proofs of the three lemmas do follow the above procedure and are shown below.

Proof. (Lemma 9) Let X_i be the indicator function that the distance between the first node and node i is smaller than $\sqrt{64n \log(n)/\pi k}$. Let $Y = \sum_{i=2}^{k+1} X_i$, i.e., Y is the number of nodes with distance to the first node smaller than $\sqrt{64n \log(n)/\pi k}$. To show this lemma, we first find out the mean of Y and show Y is around $\mathbb{E}[Y]$ with overwhelming probability. Indeed,

$$
\begin{aligned}
\mathbb{E}[Y] &= \sum_{i=2}^{k+1} \mathbb{E}[X_i] \\
&= \sum_{i=2}^{k+1} \mathbb{P}(X_i = 1) \\
&\geq 16 \log(n)
\end{aligned}
\tag{34}
$$

where the last inequality follows if the first node is located at a corner of the square.

To apply Lemma 17, we choose $c = \mathbb{E}[Y]/2$ and observe $\nu = \mathbb{E}[Y]$, and get

$$\mathbb{P}(Y \leq \mathbb{E}[Y]/2) \leq \exp(-(\mathbb{E}[Y]/2)^2/2\mathbb{E}[Y]) \leq n^{-2} \quad (35)$$

The above equation (35) implies at least $8\log(n)$ nodes close to the first node with probability $1 - n^{-2}$ and we have the lemma. ∎

Similar to the above proof, we show the rest of two lemmas.
Proof. (Lemma 10) We may assume $A \geq 10\log(n)$. Let Y_i be the indicator function that node i falls in that rectangle. Let $X = \sum_{i=1}^{n} Y_i$, i.e., X is the number of nodes falling in the rectangle. To show this lemma, we first find out the mean of X and show X is around $\mathbb{E}[X]$ with overwhelming probability. Indeed,

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{n} \mathbb{E}[Y_i] \\
&= \sum_{i=1}^{n} \mathbb{P}(Y_i = 1) \\
&\geq 10\log(n).
\end{aligned}
\quad (36)
$$

To apply Lemma 17, we choose $c = \mathbb{E}[X]$ and observe $\nu = \mathbb{E}[X]$, and get

$$\mathbb{P}(X \geq 2\mathbb{E}[X]) \leq \exp(-(\mathbb{E}[X])^2/8\mathbb{E}[X]/3) \leq n^{-2} \quad (37)$$

The above equation (37) implies at most $2A$ nodes falling in the rectangle with probability $1 - n^{-2}$ and we have the lemma. ∎

Proof. (Lemma 11) Let Y_j be the indicator function that $(i, j) \in E$. Let $X = \sum_{j \neq i} Y_j$, i.e., X is the number of one-hop neighbors of node i. To show this lemma, we first find out the mean of X and show X is around $\mathbb{E}[X]$ with overwhelming probability. Indeed,

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{j \neq i} \mathbb{E}[Y_j] \\
&= (1 + o(1))w_i.
\end{aligned}
\quad (38)
$$

To apply Lemma 17, we choose $c_1 = \mathbb{E}[X]$ for upper bound and $c_2 = \mathbb{E}[X]/2$ for lower bound. Observe $\nu = \mathbb{E}[X]$, and get

$$\mathbb{P}(X \geq 2\mathbb{E}[X]) \leq \exp(-(\mathbb{E}[X])^2/8\mathbb{E}[X]/3) \leq o(n^{-1}) \quad (39)$$

$$\mathbb{P}(X \leq \mathbb{E}[X]/2) \leq \exp(-(\mathbb{E}[X]/2)^2/2\mathbb{E}[X]) \leq o(n^{-1}) \quad (40)$$

One may observe $o(1)$ term does not affect the results. Therefore, with above equations (39) and (40), we have the lemma. ∎

B. Proofs of Theorem 4 and Theorem 5

In this section, we present our proofs for Theorem 4 and Theorem 5, the performance of our algorithm and the lower bound on the file dissemination time of any possible algorithm. We consider a random power law graph with $\beta > 2$ and nodes are only allowed to download the file from nodes at most two hops away. In Theorem 4, we set the input of Algorithm 1 as $\epsilon = 0$ and $\mathcal{L} = \infty$. Thus, nodes always request to one-hop neighbors on the social-graph. We show that our load-balancing scheme, exploiting the property social networks have small diameters, guarantees the file dissemination time scales like $\sqrt{n}$. In the proof of Theorem 5, we adopt an approach similar to that in Theorem 8 in which we find a lower bound on the transport load. We show that the performance of our algorithm only differs from the best possible file dissemination time by a factor of n^ξ for any $\xi > 0$. Our proofs are presented in the follows.

Proof. (Theorem 4) We first claim that each transmission has a rate $\Omega(1/\sqrt{n})$. To show this, consider a horizontal highway node and its corresponding stripe. Note that these n nodes are placed uniformly and independently on the square. By Lemma 10, there are $\mathcal{O}(\sqrt{n})$ nodes in this stripe with high probability. On the other hand, each node only generates at most 6 flows. Therefore, each flow through the horizontal highway node can have a rate of $\Omega(1/\sqrt{n})$. A similar argument applies to vertical highway nodes. As this is true for all highway nodes with high probability, we have the claim. In addition, there exists a constant c_1 such that each node can receive the file in $c_1\sqrt{n}F$ time slots since the transmission starts.

Similar to the proof in Theorem 7, we show the theorem by induction on k: the distance from a node to the source on the social-graph. Our claim is nodes at distance k to the source can receive the file in $c_1 k \sqrt{n} \log_2(n)F$ time slots. It is clear that the base case is true for $k = 1$. Suppose this is true for $k - 1$ and consider nodes at distance k to the source. Since each such node is not inactive at time $c_1(k - 1)\sqrt{n}\log_2(n)F$, the node must be in a binary true with an active node as the root. Therefore, this node has to wait at most $\log_2(n) - 1$ transmissions before getting served. Thus, the node can receive the file at time $c_1 k \sqrt{n} \log_2(n)F$. Hence, by mathematical induction, we have the claim. Note that the diameter of the social graph is $\mathcal{O}(\log(n))$. All nodes can get the file in $\mathcal{O}(\sqrt{n}\log^2(n)F)$ time slots. ∎

Proof. (Theorem 5) To apply Lemma 15, we just need to show that the transport load is $\Omega(n^{3/2-\xi}F)$ for any $\xi > 0$ with probability $1 - o(1)$. The idea is to show there are $\Theta(n)$ nodes in the largest component which only have small-sized 2-neighborhoods. Thus, these nodes must download the file from nodes which are geographically far away from them on the wireless-square. We first state the flow of the proof. In the first step, we claim we only need to consider a random power law graph with minimum expected degree $m = K\log(n)$ for some $K \geq 10$. More precisely, only consider nodes with weight in the region $[K\log(n)\ 2K\log(n)]$ in such graphs. We next show

that only a vanishing fraction of the number of them can find geographic proximate one-hop or two-hop neighbors. In the end, we show that $\Theta(n)$ of them are indeed in the largest component and thus, have the theorem.

We first show that we may assume that the minimum expected degree $m = K\log(n)$ for some constant $K \geq 10$. To do this, consider the original minimum expected degree $\hat{m} < 10\log(n)$ and the original expected degree sequence $\hat{w} = (\hat{w}_1, \ldots, \hat{w}_n)$. Let w be the expected degree sequence for $m = K\log(n)$ for some $K \geq 10$. Observe that (3) is an increasing function in terms of $\bar{d}$. We have $\hat{w}$ is smaller than w term by term. Thus, by coupling, the random power law graph generated by w contains the original random power law graph stochastically.

Next, we show that nodes with weight in that region have a small-sized 2-neighborhood with high probability. This property is important as small-sized neighborhood implies it is hard to find geographic proximate neighbors. Consider $2\xi > \eta > 0$. Let $\mathcal{N}_i$ be the set of nodes that node i can reach in 2 hops in the social-graph. We claim $\mathbb{P}(|\mathcal{N}_i| \leq 10Kn^\eta \log(n)) = 1 - o(1/\log(n))$. Indeed, by Lemma 11, node i has at most $4K\log(n)$ neighbors on the social-graph with probability $1 - o(n^{-1})$. Further, the probability that one of its neighbors is of weight greater than n^η is smaller than

$$\frac{2K\log(n)\int_{n^\eta}^{M} x^{1-\beta}n\,dx}{\text{vol(G)}\int_{K\log(n)}^{M} x^{-\beta}\,dx} = o(1/\log(n)). \quad (41)$$

Thus, the sum of weights of its neighbors is smaller than $4Kn^\eta \log(n)$ with probability $1 - o(1/\log(n))$. Hence, by Lemma 11 again, we have the claim.

Let $\mathcal{M}$ be the set of nodes with expected degree in the range $[K\log(n)\ 2K\log(n)]$. Then,

$$|\mathcal{M}| \approx \frac{\int_{K\log(n)}^{2K\log(n)} x^{-\beta}n\,dx}{\int_{K\log(n)}^{M} x^{-\beta}\,dx} = (1+o(1))(1-2^{1-\beta})n \quad (42)$$

We next claim only $o(n)$ nodes in $\mathcal{M}$ have geographic proximate neighbors (the distance between the neighbors and the node is smaller than $\sqrt{n/10K\pi n^\eta \log^2(n)}$). This property along with (41) implies almost all nodes in $\mathcal{M}$ do not have geographic proximate neighbors. We show this property in the follows. We first find out the expected number of nodes in $\mathcal{M}$ which have geographic proximate neighbors. Let X_i be the indicator function that the Euclidean distance from node i to $\mathcal{N}_i$ on the wireless-square is smaller than $\sqrt{n/10K\pi n^\eta \log^2(n)}$. Therefore, we have, for $i \in \mathcal{M}$ and n large enough

$$\mathbb{P}(X_i = 1) \leq 1/\log(n) + o(1/\log(n)) \leq 2/\log(n) \quad (43)$$

since the first term is the probability that $|\mathcal{N}_i| \leq 10Kn^\eta \log(n)$ and $X_i = 1$, and the second term is the probability that $|\mathcal{N}_i| > 10Kn^\eta \log(n)$. Therefore, we have

$$\mathbb{E}\left[\sum_{i\in\mathcal{M}} X_i\right] \leq 2(1-2^{1-\beta})n/\log(n) \quad (44)$$

Next, we show a concentration result. We claim that $\mathbb{P}(|\sum_{i\in\mathcal{M}} X_i - \mathbb{E}[\sum_{i\in\mathcal{M}} X_i]| \geq n/\log^{1/3}(n)) = o(1)$. Indeed, by Chebyshev's inequality, we have

$$\mathbb{P}\left(\left|\sum_{i\in\mathcal{M}} X_i - \mathbb{E}\left[\sum_{i\in\mathcal{M}} X_i\right]\right| \geq n/\log^{1/3}(n)\right)$$
$$\leq \frac{\mathbb{E}[(\sum_{i\in\mathcal{M}} X_i)^2]}{n^2/\log^{2/3}(n)}$$
$$\leq \frac{2n^2/\log(n)}{n^2/\log^{2/3}(n)} = o(1) \quad (45)$$

where the last inequality follows from $\mathbb{E}[X_iX_j] \leq \mathbb{E}[X_i]$.

At the end, let $\mathcal{S}$ be the set of nodes in the largest component. Since almost all nodes are in the largest component, we have $|\mathcal{S} \cap \mathcal{M}| = \Theta(n)$. Hence, the result follows by the above fact only $o(n)$ nodes in $\mathcal{M}$ have geographic proximate one-hop or two-hop neighbors. ∎

REFERENCES

[1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transaction on Information Theory, Vol. 46, No. 2, pp. 388-404, 2000.*

[2] M. Franceschetti, O. Dousse, D. Tse, P. Thiran, "Closing the Gap in the Capacity of Wireless Networks Via Percolation Theory," *IEEE Transaction on Information Theory, Vol. 53, No. 3, pp. 1009-1018, 2007*

[3] S. R. Kulkarni and P. Viswanath, "A deterministic approach to through-put scaling in wireless networks," *IEEE Trans. on Information Theory, Vol. 52, No. 6, pp. 1041-1049, 2004.*

[4] F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," *Internet Mathematics, Vol. 1, No. 1, pp. 91-114, 2002.*

[5] F. Chung and L. Lu, "Connected components in random graphs with given expected degree sequences," *Annals of Combinatorics, Vol. 6, pp. 125-145, 2002.*

[6] S. Milgram, "The small world problem," *Psychology Today, Vol 1, No. 1, pp. 60 V 67, 1967.*

[7] R. Albert, H. Jeong, and A. Barabási, "Diameter of the world wide web, *Nature, pp. 130-131, 1999.*

[8] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relation-ships of the Internet topology," *ACM SIG-COMM, 1999.*

[9] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Physica A, Vol. 311, pp. 590-614, 2002.*

[10] J. Grossman, P. Ion, and R. De Castro, "Facts about Erdös numbers and the collaboration graph," available from the WWW: (http://www.oakland.edu/enp/trivia/), 2003.

[11] X.-Y. Li, S.-J. Tang, and F. Ophir, "Multicast capacity for large scale wireless ad hoc networks," *ACM MobiCom 2007.*

[12] S. Li, Y. Liu, and X. -Y. Li, "Capacity of large scale wireless networks under gaussian channel model," *ACM MobiCom 2008.*

[13] S. Shakkottai, X. Liu, and R. Srikant, "The multicast capacity of ad hoc networks," *ACM MobiHoc 2007.*

[14] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks," *IEEE Trans. on Auto. Control, Vol. 37, no. 12, pp. 1936-1949, 1992.*

[15] P. Gupta and P. R. Kumar, "A system and traffic dependent adaptive routing algorithm for ad hoc networks," *Proc. IEEE 36th Conference on Decision and Control, 1997.*

[16] S. Subramanian, S. Shakkottai, and P. Gupta, "On optimal geographic routing in wireless networks with holes and non-uniform traffic," *Proc. IEEE INFOCOM 2007.*

[17] S. Subramanian, S. Shakkottai, and P. Gupta, "Optimal Geographic routing for wireless networks with near-arbitrary holes and traffic," *Proc. IEEE INFOCOM 2008.*

[18] A. Frieze and G. Grimmett, "The shortest path problem for graphs with random arc lengths," *Discrete Applied Mathematics, pp. 577, 1985.*

[19] B. Pittel, "On spreading a rumor," *SIAM Journal of Applied Mathematics, Vol. 47, No. 1, pp. 213-223, 1987.*

[20] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking, "Randomized rumor spreading," *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science, 2000, pp. 565-574.*

[21] S. Sanghavi, B. Hajek, and L. Massoulié, "Gossiping with Multiple Messages," *IEEE Trans on Information Theory, Vol. 53, No. 12, pp 4640-4654, 2007*

[22] S. Deb and M. Médard, "Algebraic gossip: A network coding approach to optimal multiple rumor mongering," *Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing, 2004.*

[23] D. Kempe, J. Kleinberg, and É Tardos, "Maximizing the spread of influence through a social network," *ACM KDD, 2003.*

[24] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," *IEEE INFOCOM, 2005.*

[25] S. Ioannidis, A. Chaintreau, and L. Massoulié, "Optimal and Scalable Distribution of Content Updates over a Mobile Social Network," *IEEE INFOCOM, 2009.*